Grant Period	5/1/2016 - 4/30/2018
Institutions	CUNY Research Foundation and The Graduate School and University Center, CUNY
Amount Awarded	\$63,485
Federal Award ID Number (FAIN)	620128194
Project Website	BeyondCitation.org
Keywords	scholarly communication, Beyond Citation, electronic resources, databases, libraries

Project Description

The Graduate Center, CUNY was awarded a Level II Digital Humanities Start-Up grant in May 2016 to further the development of Beyond Citation, a web-based research platform to facilitate the understanding of the mechanisms and functionality of commercial and scholarly research databases in the humanities. During this phase, the project developed:

- a series of essays that function as thematic guides to digital scholarship written by scholars in humanities fields such as art history, history, and philosophy, and
- a set of prototype tools for use by both research libraries and individual researchers.

The project's developers created a customized API in WordPress to allow library technologists to access Beyond Citation's content directly from computer to computer and fabricated widgets to be placed on library electronic resources webpages that will allow researchers to view Beyond Citation content describing particular databases without leaving the library's page that they are visiting.

History

The defining idea for Beyond Citation was conceptualized by Eileen Clancy¹ in the Fall 2013 Digital Praxis Seminar, a two-semester sequence at the CUNY Graduate Center, taught by Profs. Stephen Brier and Matthew Gold.

¹ The idea for the project was sparked when Clancy encountered a blog post by historian Caleb McDaniel bemoaning the lack of information about the mechanisms of commercial databases and suggesting the creation of an online repository of information. http://texas2011.thatcamp.org/04/08/increasing-proprietary-database-literacy/ and http://mcdaniel.blogs.rice.edu/?p=150

A team of graduate students created an initial pilot site for the project in the Spring 2014 semester of the Digital Humanities course. (Team members: Eileen Clancy, Rebecca Federman, Genevieve Johnson, David Naranjo, and Marisa Plumb) The project was set up as an experiment to evaluate interest among humanities faculty and librarians in an online research platform to mitigate the problem of lack of information about the mechanisms and functionality of commercial and scholarly databases.

Project Team (2015-2018)

Stephen Brier, Project Director, Professor, Ph.D. Program in Urban Education & Founder, Interactive Technology & Pedagogy Certificate Program and New Media Lab, The Graduate Center, CUNY

Eileen Clancy, Co-Project Director, M.A. student, The Graduate Center, CUNY

Boone Gorges, Lead Developer, CUNY Academic Commons, HardG

Daniel Jones, Developer, CUNY Academic Commons, HardG

Milo Axelrod, Graphic Design and User Interaction Design

Project Advisors

Jill Emery,* Associate Professor and Collection Development & Management Librarian, Portland State University

Rebecca Federman,* Electronic Resources Librarian, New York Public Library

Kathleen Fitzpatrick, Director of Digital Humanities and Professor of English at Michigan State University

Matthew K. Gold, Associate Professor of English and Digital Humanities at the Graduate Center, CUNY Graduate Center

Natalie Houston, Associate Professor of English, University of Massachusetts Lowell

Barbara Hui, Software developer, California Digital Library. Ph.D., Comparative Literature, UCLA

Hussein Keshani, Associate Professor of Art History, University of British Columbia, Okanagan

James Mussell,* Associate Professor of Victorian Literature and Director of the Centre for the Comparative History of Print, University of Leeds (UK)

Amy Papaelias, Associate Professor of Graphic Design, State University of New York at New Paltz

Claire Potter,* Professor of History and Director of the Digital Initiative, New School University

Lara Putnam,* Professor of Latin American and Caribbean history and Chair of the Department of History, University of Pittsburgh

Brian Rosenblum, Faculty Engagement Librarian for Digital Scholarship and Co-Director of the Institute for Digital Research in the Humanities, University of Kansas

Tom Scheinfeldt, Associate Professor in the Departments of Digital Media & Design and History and Director of Digital Humanities in the Digital Media Center, University of Connecticut

Polly Thistlethwaite, Professor and Chief Librarian, CUNY Graduate Center

Toni Weller,* historian of information and Visiting Research Fellow in History at DeMontfort University (UK)

*Member of Editorial Board

Project Summary

Although humanities scholars widely use academic databases from commercial publishers such as ProQuest or Gale, knowledge of how proprietary databases work is limited because their structures are dynamic and not transparent. Scholars therefore may not be aware of and cannot account for how database structures affect their interpretations of search results or text. Lack of information is an obstacle to scholarly inquiry because databases and other online resources shape the questions that can be asked and the arguments that can be made through search interfaces and algorithms.

Research databases were formerly accessible only to academics and students with institutional subscriptions. However, thousands of public libraries are now offering their patrons access with a library card. But, when college students, journalists and the public use these collections to locate information for study, fact checking, and research, they are confronted with a bewildering array of products with no roadmap to guide them. Users who may expect databases to function like Google, instead find themselves wrestling with confusing interfaces and inscrutable search mechanisms.

The Beyond Citation research platform aggregates information about proprietary commercial and nonprofit databases so that scholars can understand the significance

and provenance of the materials they glean. By making accessible essential information about the structures and content of digital resources, Beyond Citation takes an important step in updating scholarly inquiry to encourage critical thinking about these ubiquitous tools and their impact on research and scholarship. Because of the ubiquity of academic databases, we are hopeful that Beyond Citation might be consulted by scholars in every humanities discipline.

There is currently no simple, accessible way for scholars or undergraduates to understand either the affordances or the limitations of humanities databases for research. Information provided by publishers is often inconsistent and incomplete. Descriptions on library sites are usually brief summaries. Reviews in trade journals tend to be descriptive rather than critical.

Scholars' lack of familiarity with the impact of database structures on their research is at the root of the problem. Because archival databases can contain what seems to be an almost infinite number of documents, they offer an appearance of exhaustiveness that can mislead scholars and general users into thinking their search results are comprehensive. To overcome scholars' lack of knowledge about the biases inherent in databases, Beyond Citation synthesizes and codifies knowledge about academic databases, such as search capabilities, descriptions of content, provenance, legibility, and, where indicated, errors. Besides aggregating and providing links to information about digital resources, we have commissioned discipline-specific, critical thematic guides to introduce key digital database collections.

Products and scope

This NEH Level II Start Up project has four products:

- 1) a prototype research platform that aggregates information about academic databases on an open-source structure with a Creative Commons BY-NC-SA license for code and content;
- 2) discipline-specific, critical thematic essays on digital resources widely used by humanities scholars:
- 3) a custom API and
- 4) interactive, embeddable site extensions or widgets for library websites that link to Beyond Citation entries describing databases. The specially created widgets will allow us to reach database users who might not otherwise even be aware that they can find other information about the digital resources they are using.

Thirteen databases, mostly subscription-based, with a mix of publishers, types of content, and popularity, are currently included in the Beyond Citation (henceforth BC)

site, with an additional 10 sites to be added in the future. With the exception of HathiTrust, the distinguishing characteristic of these collections is their non-transparent nature.

Project Achievements

In the two and a half years since the Graduate Center was awarded the NEH Digital Level II Start-Up Grant, the Beyond Citation project has realized a number of major achievements both directly and indirectly related to the grant. The site is now running on a new infrastructure developed under the NEH grant, based on a custom-developed WordPress theme and can be accessed by librarians and technologists via an API and through widgets that can be placed on institutional webpages.

Thematic Essays

On the editorial side, we commissioned five thematic essays from humanities scholars focused on digital research relevant to their areas of scholarly expertise. These essays provide critical thematic guides to scholarly research in their disciplines which include art history, history, philosophy, and information science. These essays are posted on the project's website.

Editorial process and decisions

In December 2016, we held a virtual brainstorming session between BC project staff and members of the project's editorial board (drawn from our Advisory Board). Editorial board members shared their recommendations for developing a framework for providing style and substantive guidance to scholars who would be writing critical thematic guides to specific electronic resources.

The editorial board discussed the style and general approach of the essays and the editorial process that we would need to implement, including length of the essays (2,000-5,000 words), tone, citation styles, the review process, the selection of outside reviewers, and the incorporation of post-publication changes, revisions and comments.

Considerable discussion ensued among advisors and BC staff about how to combine formal academic rigor with accessibility and readability for BC's anticipated broad audience. We decided, given that we were planning to critique the obscurantist nature of many electronic resources, that clarity of expression for the writing on the website would be crucial. It was agreed that level of writing on the site should be geared to an audience of motivated undergraduates.

Editorial board members felt strongly that the essays should take advantage of the digital affordances of the web where possible and that we should encourage writers to work with BC staff to incorporate visuals into the online presentation of the essays on

the website. Besides the visual aspects of the essays, advisors also raised the question of how we might best be able to incorporate comments, corrections and updates to the published essays.

We believe it is very important that any changes made to the website should be fully citable. We talked about version control for the essays (and the website), something that is both a technical and an intellectual challenge. How should we make changes visible publicly? How might we include comments and changes to essays that describe and critique digital resources that are themselves dynamic? One advisor suggested that publication of the critical essays be seen as a fixed starting point and that we imagine revised, annotated, and updated versions as "new editions." In these early discussions, we agreed that it was not possible to make final decisions about the best ways to reflect changes in the website because that process would be highly dependent on the final technical architecture of the website. We agreed that it would be important to return to the subject later on and to bring our designer and user experience (UX) advisors into future conversations to help think about the relationship between design, navigation and citation.

For the editorial review process, we decided to implement a structured evaluation process aiming primarily to improve the essays and increase their accessibility for a more general audience. With that in mind, the project director and project coordinator decided they would, as a first step, read all essays and quickly offer initial comments and suggestions to writers when the essays were submitted. Thereafter, the revised versions were circulated, in some cases to individual members of the editorial board, who returned comments to BC staff if they wished as well as offering suggestions about possible additional outside reviewers.

Because we were committed to the idea of open peer review, the essay writers were informed of the identities of the outside reviewers and asked if they were aware of any reason that the reviewer might not be appropriate (for example, if there was a conflict of interest, or a close prior personal relationship). In one instance, a reviewer asked for anonymity, a request that we honored. But almost all reviewers were comfortable with the idea of their identities being transparent to the essay writers.

Our goal in this process was that formal outside reviews of each essay would be performed by two people: a specialist in the scholarly field covered by the essay and another reviewer who was a generalist in the field and who could make suggestions to improve the accessibility of the essay for readers outside of the essay writer's academic discipline. In some instances, Beyond Citation advisors played either of these reviewing roles. In the end, every essay had at least one outside reviewer; one had three.

Editorial guidelines

Based on these conversations with the editorial board, the project director and project coordinator created a formal set of essay drafting instructions (See Appendix A) for prospective essay writers. In creating these, the project coordinator consulted the style guidelines posted by *Kairos: A Journal of Rhetoric, Technology, and Pedagogy* (and former NEH-ODH grant recipient) and the digital humanities *Journal of Interactive Technology and Pedagogy*.

Essays commissioned and published online

"Digital Resources in Modern Chinese History"
Maggie Greene, Assistant Professor of History, Montana State University, Bozeman

"Ancestry.com: From Public Records to Genealogy Big Data"
Katharina Hering, historian and archivist based in Washington, DC. Ph.D. in American History, George Mason University; MLIS, University of Pittsburgh

"Digitizing Islamic Architectural Memory: A Review of Archnet 2.0" Hussein Keshani, Associate Professor of Art History, University of British Columbia, Okanagan

"LGBTQ Archives at the Schlesinger Library and the Sophia Smith Library" Claire Potter, Professor of History and Director of the Digital Initiative, New School University

"Stanford Encyclopedia of Philosophy: A Review"
Chris Alen Sula, Associate Professor at Pratt Institute's School of Information and Coordinator of Digital Humanities and Program Coordinator of Data Analytics & Visualization, School of Information, Pratt Institute

Two essays replaced essays in the initial list proposed in original NEH grant application. The first one on Ancestry.com replaces a proposed essay on U.S. government databases; the second, a review of the *Stanford Encyclopedia of Philosophy*, is a new essay not contemplated in the original grant proposal to the NEH.

[Because of family obligations, our advisor Natalie Houston was not able to contribute an essay on 19th Century English literature, as originally envisioned in the proposal. We are hopeful that she may contribute an essay on this topic at a future date.]

Design process

In Spring 2018, the BC project staff hired a visual designer, Milo Axelrod, based in New Paltz, New York, whose principal role has been to implement new pages contemplated for the site, especially the space for the presentation of scholarly essays, several of which are more than 7,000 words in length.

Axelrod (who uses they, them pronouns) created a customized space for the essays; new pages (a Colophon page and a page for API documentation). They implemented new navigation; performed dozens of small tweaks; and smoothed out issues related to the display of text with the new website plugin. The space for reading essays is a simple, uncluttered interface well suited for the consumption of long-form pieces. It has the affordances of a simple, mobile-friendly platform and, at the same time, offers easy navigation to allow the user to go back and forth between the footnotes and main text.

With an eye toward easing maintenance tasks on the website, at the request of the project coordinator, the designer also created a space within the WordPress dashboard for documentation to be consulted by BC staff and volunteers. The documentation offers detailed instructions such as how to create footnotes in the essays or the proper way to add images so that the site maintains its responsiveness.

Communication Among Team Members

While the work of some digital humanities projects is gathered around a core group in a digital humanities center or physical location in a library, except for the core team, people connected to Beyond Citation were widely dispersed. The core team exchanged regular emails and at various junctures in the process met as often as every week and at other times only every couple of months. But nearly all of the other work on the project took place remotely. The programmers worked remotely and have never met the project coordinator in person. The designer has yet to meet any BC staff in person and only a few members of the editorial board have ever met the staff, or one another for that matter. Many tech companies now have a large percentage of their staff working remotely as new affordances have been developed to make long-distance work viable. Even large companies, such as GitHub, now have their entire staff working remotely. That said, it is necessary to consciously find ways for all of collaborators and advisors to be productively connected to the project and to lend their unique talents and expertise to it.

To initiate and maintain the strength of those connections, even though most of the people contributing to the project are technically not working for Beyond Citation, the project coordinator used an onboarding type of process with many of them. The project coordinator has gone to considerable lengths (attending conferences in multiple cities)

to speak with the essay writers and advisors in person about the BC project. She has also spoken to almost every advisor on the phone.

As far as the digital tools that we employed old-fashioned email, of course, is often the glue holding communications together. For synchronous conversations among large groups, we used video conferencing. The design work on the project has been largely performed through collaborative consultations between the designer and the project coordinator talking on the phone, messaging on a Slack channel and tracking tasks and milestones with the online project management tool Notion.

Overall, these systems worked well and communication was largely untroubled.

Technical Issues

The key focus of our programming work was augmenting the usefulness of the "Explore the Databases" section of the site, which offers detailed descriptions of specific electronic resources. In that section of the site, we present a collection of information curated by BC staff including pertinent factual descriptions, information about provenance and access, and scholarly writing on the resources. The central content is high-level descriptions of databases and links to titles lists. Each entry describing a database is a single page with a visual display of tabs (see screenshots below as well as Appendix B, which details the underlying BC data structure).

Explore the Databases

Google Books

The Google Books Library Project is digitizing millions of books and magazines from major research libraries such as the University of Michigan, the New York Public Library, and Oxford University. Google Books searches the full text of these whether they are in copyright or in the public domain.

Nineteenth Century Collections Online (NCCO)

Nineteenth Century Collections Online (NCCO) is a curated compilation of primary source documents gathered and digitized from archives from around the world, focusing on eight major themes of the 19th century: Asia and the West;

ProQuest Historical Newspapers

ProQuest Historical Newspapers includes the full run of several major U.S., international, African-American, and Jewish-American newspapers. The content includes the full image of the article as it appeared on microfilm, including advertisements.

JSTOR

JSTOR is a multi-disciplinary journal archive and a platform for scholarly ebooks and primary sources. Archived journals date from the earliest issue published. Journals in the JSTOR Current Scholarship Collection include the most recent issues without an embargo.

Project MUSE

Project MUSE is a humanities and social sciences scholarly journal and ebook platform. Project Muse includes journals and books from major academic publishers in the United States and many from around the world.

19th Century U.S. Newspapers

19th Century U.S. Newspapers is a full-text, facsimile image database of approximately 500 national and regional newspapers, dating from 1800-1900.

HathiTrust Digital Library



About V Explore Databases Essays For Librarians V Blog

Google Books

march 20, 2018

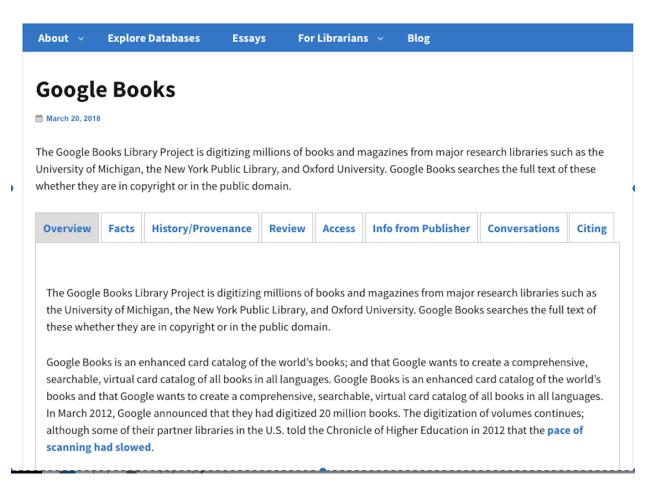
The Google Books Library Project is digitizing millions of books and magazines from major research libraries such as the University of Michigan, the New York Public Library, and Oxford University. Google Books searches the full text of these whether they are in copyright or in the public domain.

Overview Facts History/Provenance Review Access Info from Publisher Conversations Citing

The Google Books Library Project is digitizing millions of books and magazines from major research libraries such as the University of Michigan, the New York Public Library, and Oxford University. Google Books searches the full text of these whether they are in copyright or in the public domain.

Google Books is an enhanced card catalog of the world's books; and that Google wants to create a comprehensive, searchable, virtual card catalog of all books in all languages. Google Books is an enhanced card catalog of the world's books and that Google wants to create a comprehensive, searchable, virtual card catalog of all books in all languages. In March 2012, Google announced that they had digitized 20 million books. The digitization of volumes continues; although some of their partner libraries in the U.S. told the Chronicle of Higher Education in 2012 that the pace of scanning had slowed.





The project faced unanticipated technical challenges in re-purposing a WordPress installation to present and offer API access for structured data. In the initial site, the visual display of tabs was created by jQuery shortcodes that are trivial to implement. This functionality was essentially an editorial rather than a code-based construction. The content of this section is narrative information that was stored on the site in text blocks and generated from a blog page in WordPress. The visitor's movements through the website text were triggered through interaction with visual tabs. While the "Explore the Databases" section on the website looked as if its information had been generated from

fields in a database, in fact, at the source code level, it was simply a compilation of unstructured data on a standard WordPress blog page.

Converting a blog into a website containing structured data, with a usable API for library technologists to engage with would prove to be a challenge that was not possible to overcome while staying within bounds of the project described in the NEH grant proposal and staying within the proposed budget.

In order to understand why this issue proved so critical for the project, we need to revisit the project's earlier history. Prior to receiving the Start-Up grant, and during early discussions about how we might encourage library patrons to find and use information in BC, Rebecca Federman, Electronic Resources Librarian at the New York Public Library and a BC advisor, suggested that a widget might be a viable solution. One of the key advantages of a widget is that it does not require a user to exit the library's website to get to the resource; the information is generated on the same page. However, to power the widget, we would need to have the capability of pulling data from the WordPress API.

We believed that widgets would be crucial to generate interest in the project within the library community. Thus, in order to encourage libraries to offer a simple way to connect Beyond Citation with their patrons, the project staff concluded that we would have to craft a widget and customize the WordPress API. Adding this functionality to the project, was a key aspect of our grant application to NEH-ODH.

Phase 1 of Programming

Technical consultations—Summer-Fall 2016

After receiving the NEH grant, and before starting to build, we continued pursuing technical advice and consulting anew with in-house technology experts and outside programming advisors regarding the choice of and construction of the digital platform for BC. We decided to be open to considering a variety of digital platforms in order to determine which might be most efficacious for realizing the project's goals. We thought about sticking with WordPress (our first choice); using Omeka; and even Django/Flask, pondering how each might fit with our requirements. At every turn, we came back to WordPress as the most suitable CMS.

Before a final decision was made as to the choice of platform for the project, during fall 2016, discussions began with staff and developers from the CUNY Academic Commons (CAC) group. The personnel involved were: Daniel Jones, CAC Assistant Developer and development lead for the BC project; Boone Gorges, CAC Lead Developer, serving as a

technical advisor overseeing the project, and Professor Matthew Gold, the director of CAC and a BC advisory board member.

We engaged in preliminary conversations with CAC personnel as to how our initial vision for the website might be implemented online, discussing these and other topics: the desired functionality of the site, the widgets and the API; the overall scope of our specific request to the programmers; communication among and between project staff and the development team; and project timeline and deadlines.

In the process, we encountered some difficulties of translation between spheres of expertise. We were surprised to discover that WordPress developers have very different ideas about the functionality of widgets than library technologists do. We worked hard during Fall 2016 to overcome these gaps in our team's knowledge and in our ability to explain the functionality we wanted for the BC site in language recognizable to programmers.

Project kickoff meeting-December 2016

The project kickoff meeting on the programming side took place in December 2016 during which the BC project team and team developers reached a shared understanding of the work requirements. That understanding was memorialized in a letter after the meeting.

[NB: The schedule for the implementation of the first work plan for Beyond Citation had to be pushed back by three months as a result of the Project Coordinator's medical problems.]

As mentioned earlier, it was necessary to take the preexisting content that had been developed before the NEH grant and that had been stored in WordPress in text blocks and create a new schema to organize the information. The programmers then had to rewrite the existing WordPress theme to pull content from the appropriate location.

The project coordinator supplied an initial schema for the "Explore the Databases" data on the site and shared it with the programmers. BC developers created a document describing the existing tab/field mapping, noting discrepancies that may have existed between different databases (outliers in the descriptions). This document was presented to the BC team for verification and to suggest modifications necessary to encompass all databases in the Explore Databases section of the BC website.

With this map in hand, BC developers created the plan for the custom field structure. They then rewrote the existing WordPress theme to pull content dynamically from the

structured data. On individual database pages, they also built tabbed navigation into the theme plugin, rather than having it provided by a shortcodes plugin using jQuery. (A further benefit of this method is that online resources such as the Internet Archive are now able to capture a far more accurate view of the BC site than can be done on a site that uses jQuery which is very difficult for current web archiving tools to emulate correctly.)

The development process took place in a "sandbox" (demonstration or trial) space to allow the project team to evaluate the capability of the platform to handle the structured data. After the widget generator was completed, Daniel Jones and Boone Gorges then developed a Databases API plugin for BC.

Testing-Spring 2017

In Spring 2017, the programming team delivered the first version of a widget generator interface and an API plug-in for the WordPress theme. We took the initial prototype plugin that the developers had created and shared access to the sandbox space for testing with Graduate Center Library Digital Services Librarian Stephen Klein and Data and Digital Projects Librarian and programmer, Steven Zweibel, as well as Graduate Center Digital Fellow and doctoral student and programmer Patrick Smyth.

The BC staff and our technical advisors found that the tools the programmers created worked along the lines that we had anticipated; But we were unable to realize the functionality needed within the original WordPress framework. What happened?

During testing, the widget generator interface and API plug-in for the BC WordPress theme performed adequately to accomplish their immediate tasks. However, during the testing process, we realized that a problem that had been raised in our December 2016 discussions about programming would now have to be addressed by the development team.

Dual Entry Problem

A workflow problem emerged that, if it were not fixed, would have significantly handicapped the process of inputting data as well as the process of performing regular and repeated editing and maintenance on the site.

The issue we encountered arose because of the way information is stored in WordPress. In order to make the API work properly, it was necessary that information about databases—fields such as "Publisher" and "Original sources"—be stored discretely. This configuration was incompatible with the setup of the existing website, where information about a given database was stored in a single block of text, differentiated only by

headers. The result was a situation where information describing the **databases on the website had to be entered and maintained in two separate places:** in the structured schema that powers the API, and also in the text blocks that powered the beyondcitation.org graphical website.

This was a major stumbling block. The requirement that all data entry had to be exactly copied and pasted information into two separate places within the project's WordPress dashboard was an obvious major area of potential problems, if not outright failure, at a future date.

We chose WordPress as a platform for the project because it has affordances that allow non-experts to enter information into its dashboard and to perform regular updates. (This requirement is typical of many other digital humanities projects based in educational institutions and reliant on student and/or voluntary unpaid labor.) Yet what we were trying to do in with WordPress, a blogging platform, that is designed merely to display information in a browser, was to have it function itself as a database to power the sharing of structured information via an API, a task for which the platform was not designed.

During the summer of 2017, we continued to discuss with Graduate Center staff a variety of options that would mitigate these issues. We spoke about the pros and cons of attempting a non-WordPress solution to the problem, even going so far as to create a dummy prototype with Django/Flask. But every time we considered alternative platforms, we were drawn back to WordPress because of its ease of use for data entry, its huge pool of expert programmers and extensive documentation. If we were to choose not to undertake the necessary additional programming work that would be necessary, but which had not been originally budgeted for, the result would be continuing major problems maintaining the site because of the need for dual entry of information.

It became clear that in order to make the project viable, we were going to have to find supplementary money to pay for programming work not anticipated in the original NEH grant award. And, besides the financial cost of adding an additional stage of programming, there was the additional cost in time and labor as well. The new plan for completion of the project would necessarily diverge from our originally proposed schedule and would require a meaningful amount of unscheduled BC staff labor to oversee and implement programming, as well as changing the editorial schedule.

During these discussions, and indeed, in all phases of development, project advisor Prof. Matthew Gold played an essential role. The project benefited as well from the counsel of Prof. Lisa Rhody, Co-Director of the Graduate Center Digital Initiatives program, who

has extensive experience as a project manager, including specifically in the WordPress environment.

Project staff then asked the programmers to develop a proposal to ameliorate the dual entry problem as outlined by the staff. The development team estimated we would need an additional \$7,500 of programming time (an increase of 75% over the budget allocated for programming in the original proposal to NEH-ODH) to implement the fix they suggested—the creation of a new theme.

At this point, Prof. Gold stepped in and offered to contribute an additional \$5,000 to BC out of the Graduate Center's Digital Humanities Initiative budget. Thankfully, with this additional financial support, along with the decision by the project director to forego his \$4,000 summer salary on the BC grant as well as several of our project advisors who waived their consulting fees, the project was assured of the necessary financial resources and was able to move forward. And so we extended our work plan for programming and testing into the future and began Phase 2 of the project.

Phase 2 of programming (August 2017–April 2018)

The second phase of the Beyond Citation project involved the reworking of the http://beyondcitation.org theme to eliminate the "dual entry" problem described above. The BC team provided a new information schema and Boone Gorges rewrote the existing theme to pull the content as well as made the necessary modifications to the BC API plug-in.

By the end of 2017, a second version of the Beyond Citation WordPress theme had been created by developers and shared with staff and consultants within the Graduate Center. Boone Gorges configured and hosted a clone of the existing beyondcitation.org website on the website of his company HardG. The BC team used the clone site for periodic testing. Iterative work and testing continued with feedback to the programmers from BC staff over the next six months. During this process, the project coordinator evaluated the performance of the plugin in relationship to the desired display of data on the BC website, requesting numerous minor tweaks that Gorges implemented.

Toward the end of this period, the project coordinator migrated all of the existing data in "Explore the Databases" into the theme being built in the project sandbox. In April 2018, the project director installed the new plugin on the BC site. The process was successful and resulted in all functionality and display occurring as the staff expected.

Widget Generator

Beyond Citation provides HTML and JavaScript code that automatically pulls information about the selected database from the BC API, and displays it on a library website. Each discrete digital collection requires a separate set of code. To get the widget code for a particular website, users would navigate to the database they are interested in and click the Embed button.

Home About FAQ Sample Page Team

Early English Books Online (EEBO)

October 25, 2017

EEBO contains over 100,000 early English book titles, including incunabula, reflected in bibliographies such as the **English Short Title Catalogue**, Pollard & Redgrave's Short-Title Catalogue and Wing's Short-Title Catalogue.



Two buttons appear—"Dynamic" or "HTML-only." When users click either button, a line of code will appear that they can copy and paste into their website. [Mock-up shown does not display actual code.]

Home About FAQ Sample Page Team

Early English Books Online (EEBO)

October 25, 2017

EEBO contains over 100,000 early English book titles, including incunabula, reflected in bibliographies such as the **English Short Title Catalogue**, Pollard & Redgrave's Short-Title Catalogue and Wing's Short-Title Catalogue.



Copy the HTML below and paste it into your website to add a widget with information about Early English Books Online (EEBO). Click anywhere inside the text area to to select it for copying.

<pre><div class="bc-widget" data-database="early-english-books-online-eebo"></div> <script src="http://boone.cool/beyondcitation/wp-content/plugins/beyond- citation/widget.js" type="text/javascript"></script></pre>							
Overview	Facts	History/Provenance	Review	Access	Info from Publisher	Conversations	Citing

Outreach

In January 2018, the project coordinator Eileen Clancy took part in a roundtable discussion co-organized by the National Historical Publications and Records Commission (NHPRC) of the U.S. National Archives and Records Administration and the American Historical Association (AHA). This session was part of a series of seminars, "Primary Sources and the Historical Profession in the Age of Text Search." The session she participated in was called "Historical Research and Analysis in the Digital Age." [https://aha.confex.com/aha/2018/webprogram/Session16401.html] The panel discussion was framed around the "new ways historians are now working and accessing primary sources, as well the need for a more active involvement of historians in the development of this emerging historical infrastructure." The panelists addressed the question: "How can individual scholars, departments, libraries, and institutions collaborate more fully to maximize the potential of digital primary sources for historical research?"

In the final roundtable of this same AHA series during the 2018 annual meeting in Washington D.C., historians and BC advisors Lara Putnam (University of Pittsburgh)

Beyond Citation White Paper,
page 19 of 29

and Hussein Keshani (University of British Columbia Okanagan) took part in a session on "Theory and Method in the Digital Age."

Both sets of sessions were covered in a lengthy article by Stephanie Kingsley Brooks in AHA's *Perspectives* magazine [https://www.historians.org/publications-and-directories/perspectives-on-history/march-2018/search-history-making-research-transparent-in-the-digital-age] that quoted the three BC representatives at length. Clancy spoke about how we can learn from book historians and apply their ideas about the instability of textual transmission to digital resources. Keshani pointed to the "creation of a digital surrogate" as a "vulnerable moment." Putnam warned that the historian's craft was under threat when they were working in digital realms without understanding their sources or processes.

Launch

The project has been undergoing a "soft launch" during the spring and summer of 2018 during which time the website infrastructure has been updated, design features have been smoothed out, and the five commissioned essays have been posted. In this period, our advisors are offering comments on site navigation and related issues. Programming staff has also been making minor changes to the plugin. The website will be formally launched with a public announcement in early September 2018, with outreach on social media. We are planning to hold a public launch event at the CUNY Graduate Center during the Fall 2018 semester to formally announce the Beyond Citation site and to encourage academics and the general public to begin using the site and to offer their feedback on its functionality and utility.

Lessons Learned

Many digital humanities projects take place within turnkey centers or in institutional libraries. But, with the exception of the project director and the project coordinator who are New York City-based, key team members in this project live in different cities. Additionally, the large number of advisors is a far-flung group, living across eight time zones, from Vancouver, B.C. to England. Having such a big group is wonderful for providing a variety of perspectives, and perhaps even essential for such a multidisciplinary project. Nonetheless, it proved difficult to keep the communication going smoothly between all parties.

Considering this was a first attempt at creating new editorial content, commissioning so many essays was overly ambitious, particularly in what Hussein Keshani has described as writing in a new genre: the critique of digital scholarly resources. Essentially, there is no comparable publication, online or otherwise, that undertakes full-bore reviews of library-based digital resources for scholars. When essay writers had questions, we had

no models to point them to. That the essay writers were from so many different academic disciplines was quite challenging as well, both for identifying appropriate peer reviewers and to stretch our own understanding of their diverse academic fields. And, getting academics to write in a way that is accessible to a fairly general audience (even when that audience includes other scholars) is tough! The editing process was extremely time-consuming.

Additionally, most of the editing and peer review process took place concurrently with the unplanned-for second phase of programming. With only two members, the project staff was forced to engage in a sort of staccato process that unfolded in fits and starts.

The key problem we encountered that proved frustrating for our production process was the need to use the WordPress CMS as the de facto substrate for a digital humanities project regardless of its suitability, and having to develop cumbersome workarounds. This situation is not going to go away for us or for other DH projects. There is not an obvious solution, when one is creating public-facing projects, largely because creating a new product/CMS would require a very large investment of funds and, for success, would require attracting a massive number of developers and users into its ecosystem.

Furthermore, both the project director and the project coordinator were unable to work on the project for medical reasons for various and different stretches of time during the period of the project, which necessitated a request to NEH for a one-year no-cost extension and caused significant delays in the timing of our announcement of a relaunched website.

Having the timeline of this project extend over a two and a half year period caused other unanticipated effects. The project coordinator, who was tasked with handling the majority of the day-to-day needs of the project, is a Masters student without fellowship support who has had to work at a variety of other jobs throughout the project period. Inevitably, at times, project work would be delayed because the project coordinator was making a living elsewhere. This circumstance affected the forward momentum of the project on multiple occasions. Additionally, to an extent that exceeded our initial expectations of his workload, the project director had to step up many times to take up the burden of administrative work and, on the editorial side, communication with the essay writers and peer reviewers. In brief, the funding for staffing for this project fell short of what it should have been given the project's intellectual ambitions.

Further Work

We are committed moving forward to expanding the current functionality of the existing project in a variety of areas.

We would like to analyze and add more coverage of database resources in the "Explore the Databases" section of the site. Once the schema has been constructed, this type of work can be parceled out among many hands. Our hope is that librarians will be interested in engaging with that project. We are in touch with at least one information school professor who has expressed interest in having us work with their students to build out additional database entries. And we have friendly relationships with key librarians involved in critical librarianship (for example, the editor of Library Juice Press, and the Venn Diagram of people who convene for regular social media chats under the hashtag #critlib, and Beyond Citation's followers on Twitter.)

We have a strong relationship with the leaders of the American Historical Association, who have articulated the goal of getting historians to embed digital methods in their work and to learn to interrogate digital resources in meaningful ways. The AHA's director of digital projects, Seth Denbo, participated in a scholarly communications panel on historians' use of digital resources organized by project coordinator Eileen Clancy at the 2015 Advancing Research Communication & Scholarship conference. And, most recently, Clancy and two BC advisors participated in a series of roundtables at the 2018 AHA conference, described in the Outreach section above.

The Center for Research Libraries (CRL) has been a strong supporter of the project, with CRL President Bernard Reilly writing a letter of support for our NEH-ODH application. We hope to be able to work with CRL to reach out to their 200-strong member libraries to encourage them to use Beyond Citation and to incorporate widgets on their library pages.

BC staff plan to pursue additional funding in the near future to carry the project forward.

Conclusion

Digital tools for access to repositories of knowledge in the humanities can and must do much more work than act merely as file folders for archival documents and artifacts. Search algorithms and controlled vocabularies perform invisible labor that affects our ability to access knowledge on a daily basis.

There has been surprisingly little consciousness, until quite recently, among traditional scholars around the issue of interrogating digital collections in the same way that a trained academic interrogates any processed physical archival collection. However, a new consciousness has recently developed in the academy around the importance of understanding the role of the digital realm in the public sphere and its impact on politics, elections, ethnic minorities, migration, war and other issues. These changes give rise to new possibilities for reaching the humanities scholars in our target audience who are

increasingly concerned with the lack of transparency in accessing digital resources and utilizing digital methods.

And, since this project was first conceptualized, change continues apace.

Digitally based knowledge tools that did not exist a few years ago now have an extraordinary impact globally; scholars must keep up with these trends. Wikipedia was founded only in 2001. At the time, it was mercilessly mocked by academics. Later on, as the site was built out, Wikipedia was criticized initially for occluding the role of women in history and for its Anglophone bias. Yet Wikipedia has increasingly taken on the role that was formerly inhabited by the research article. Now, because all of its articles require citations, Wikipedia is one of the top DOI referrers in the world with a reach far exceeding academic metrics such as the Impact Factor or rankings on Google Scholar or Scopus.

As more governments get the hang of digital communications and technology, some have ventured into areas formerly known as mostly the province of scholars, making efforts to shape ideas about the past in the hope of controlling the present and the future. For example, the Chinese government is now censoring subscription databases containing historical documents and artifacts in real time. Maggie Greene, who wrote the BC essay on Chinese online databases, has called this process, in an email to BC staff, one of "insta-redaction." Scholars need to be aware of both deliberate machinations to massage sources of knowledge in the hope of creating new realities and the hapless errors or biases that creep into any type of technological system.

We hope that the collection of essays that we have published and the larger website itself will help to create a space to address concerns about obstacles that the digital realm presents to scholarship. Some will be recognized as similar across fields and disciplines; but each must be understood at the granular level in order to be overcome. We seek to jumpstart conversations among academics and other serious researchers, particularly in the fields of librarianship, scholarly communications and in the disciplines that the essays address.

The historian T.J. Stiles has a simple rule that we think should guide future scholarly inquiry in digital spaces. "Know why you know something."

As limited in scale as this project may be, considering the tens of thousands of commercial databases being offered to institutions for scholarly use, it is hoped that the questions that we have raised will spark new discussions, new projects, and approaches.

Acknowledgments

Project Director Stephen Brier provided intellectual and strategic leadership and oversaw all aspects of the project as well as representing the project to the Graduate Center administration and handling our relationship with the Research Foundation of the CUNY, which has responsibility to oversee all CUNY the grant funded projects. He also worked closely with key editorial board members, recruited peer reviewers and several essay writers, and edited all of the submitted content essays. Brier edited the Interim Reports, White Paper, and Final Report as well. Project Coordinator Eileen Clancy edited the site content, oversaw the development for the project and relationship with programmers, acted as managing editor for the essays, edited all of the submitted content essays, wrote the Interim Reports, White Paper, and Final Report for NEH-ODH, recruited peer reviewers, maintained relationships with Advisory Board members, engaged in outreach, partnered with designers to create new spaces on the site and the logo, and maintained the platform.

The strength of the project's relationship with programmers Boone Gorges and Daniel Jones of the CUNY Academic Commons team was a bright spot throughout the period of the grant. With Gorges providing technical project management and oversight, Jones built the initial custom WordPress plugin and widget generator. Gorges followed up in the second phase of programming, crafting another edition of the plugin and multiple versions of the custom Beyond Citation theme. Gorges's patient approach to communication managed to offer just enough instruction to allow a project coordinator, who is still learning her way around a server, to successfully tackle occasionally daunting back-end tasks while always knowing that a safety net existed. It would be difficult, and likely impossible, to find a partner more knowledgeable about programming in the WordPress ecosystem with the graciousness to wear his expertise so lightly as Boone Gorges.

Our advisor Amy Papaelias took on the task of moving a website that telegraphed the visual message that it had been built by students as a class project. She created a striking and modern logo for the project as well as refreshing the appearance of the entire site, including typography and navigation. Papaelias has been a wonderful and patient partner whose knowledge of design and user interaction underlies every good and attractive aspect of the website.

We are grateful too that she referred visual and Interaction Designer Milo Axelrod to us. With great flair and in a spirit of bonhomie, Axelrod created the customized space for the essays on the website; new pages, implemented changes to navigation and smoothed out many issues. It has been our good fortune to work with them and we look forward to a continuing relationship.

Ongoing discussions with the staff of the Graduate Center's Mina Rees Library have been essential to the project's success from its earliest conceptualization. Chief Librarian Polly Thistlethwaite has unstintingly lent the BC project her considerable intellectual heft and her knowledge of the library world. Digital Services Librarian Stephen Klein and Data and Digital Projects Librarian and programmer Steven Zweibel have been extraordinarily generous with their time (often requested in the midst of crises) sharing their observations and unvarnished opinions with the staff. Jill Cirasella, Associate Librarian for Public Services & Scholarly Communication, has been a stalwart supporter and friend to the project. We're deeply grateful for their support without which we would never have achieved our goals.

We are fortunate to have had the support and guidance of the Graduate Center Digital Humanities Initiative (GCDHI) led by Prof. Matthew Gold and Prof. Lisa Rhody, which also provided supplementary fiscal support to the Beyond Citation project. Additionally, in his role as GC Digital Fellow, Patrick Smyth has shared his technical expertise, his sangfroid attitude as well as acting as an indispensable sounding board at every stage.

The fifteen members of our Advisory Board who represent many different humanities field and three countries have given the project the intellectual depth necessary to take on such an ambitious multi-disciplinary project. We owe a special debt to Prof. Amy Papaelias who created our logo, oversaw a visual refresh of the site, and referred graphic and user interaction designer Milo Axelrod to us.

We are deeply grateful for the support of scholars Jennifer Guiliano, Ian Milligan, and Michael Widner, who, at multiple junctures in the grant application process, offered comments on grant proposals, practical advice, and unabashed encouragement on the nascent project idea without which the project may well have faltered.

Dan Cohen, formerly of the Digital Public Library of America, and Bernard F. Reilly of the Center for Research Libraries both wrote exceptionally powerful letters of support for our grant application detailing their views of the necessity of a resource like Beyond Citation.

We are profoundly appreciative of the scholars who participated without compensation in the peer review process for the essays on the site: Jennifer Guiliano, Susan Kriete, Robin James, Rachel Mattson, Ruth Mostern, Lara Putnam, and members of the BC editorial board.

Beyond Citation: Critical Thinking About Digital Research NEH Digital Humanities Start-Up Grant: White Paper 2018

We thank Caleb McDaniel for writing a 2011 *THATCamp Texas* blog post about historians' research practices suggesting the creation of an online repository of information about proprietary databases and for his generosity in sharing his idea with a student and the broader digital humanities community.

All of the NEH-ODH staff members were extraordinarily helpful throughout the application process and thereafter. But, we must acknowledge our program officer in particular. Over the two years that we were engaged in the application process with NEH, Perry Collins answered every query within the hour, whether during a weekday or weekend. It was only on the strength or her comments on early drafts of our application, that we had a hope of improving the application sufficiently to receive an award for this Start-Up grant. We are incredibly thankful for her intelligence and diligence.

Appendix A: Editorial guidelines

Beyond Citation essay writers' guidelines - 2018

Our overall goal in this process is to have your essays contribute to a demystification of electronic resources for both academic and general users. With that in mind, we have created the following writing guidelines for you to consider as you shape your essays for Beyond Citation.

Length: We anticipate that the essays should run between 2,000 and 5,000 words. This is flexible, of course. As you are writing, if you anticipate a significant divergence from the lower or upper levels of essay length, please let us know right away.

Tone: The tone of the essays should be somewhat formal, while being accessible and understandable to an imagined reader who is a motivated undergraduate. Because we are engaged in critiquing electronic resources that are frequently obscurantist, clarity of expression is crucial for all writing associated with the project.

Critical edge: We are asking you to draw upon your expert knowledge to write an essay with a strong, critical edge. The essays should be written without unnecessary jargon and, if warranted, with parenthetical explanations so that they may be understood by readers who are not specialists in your discipline. In a period when the proliferation of fake news and misinformation is rampant, we want to showcase models of evidence-based critical thinking, rather than voicing claims supported by authoritative credentials and cultural capital alone. Your essay should therefore be persuasive and evidence-based, displaying particularized knowledge that is sufficient to clearly and demonstrably support your arguments, pro and con, about the database(s) under review.

Based partly on your choice of databases to review in your essay, we may add additional entries (placed separately on the website) as necessary or appropriate. We will of course let you know what other databases will be reviewed in your area of expertise.

What to include: For consistency across the essays and in order that they complement our own short descriptions of the databases on the website, we ask that you consider including the following topics in the essays you are writing.

Types/varieties of materials/texts/artifacts included in the database(s): be specific and detailed. For example: manuscripts, codices, court records, photographs, maps, diaries, census documents, personal correspondence, posters, films.

Omissions: How representative across a variety of categories is the database? Major omissions in a particular digital collection may result in it being unrepresentative in a meaningful way leading to misimpressions on the part of researchers or encouraging false confidence about its completeness. The claim that a database of primary documents covers a particular time period or geographic region should be investigated to the extent that is reasonably possible. For instance, if a database of nineteenth-century documents and artifacts advertises that it is complete for that historical period, essay writers should directly interrogate the claim and try to determine (where possible) the percentage of documents authored by or concerned with women, members of subordinate classes, racial and ethnic minority populations, non-dominant languages, and so on.

There may also be other types of omissions of a more technical nature, such as when certain years are missing from the sequence of a print run of newspapers or when photographs are not included for copyright reasons. To the extent these omissions are known or clear, these types of gaps should be mentioned.

Provenance of materials/construction: For databases, as with any archive, understanding the provenance of materials is key to sense making about the material in question. In many instances, information about the construction of electronic resources may be difficult to locate or skimpy. To the extent that you are able to learn about the process by which the database was made, it would be valuable to include such information in the essay. This is an area that Beyond Citation staff would be pleased to assist you with and should be part of the ongoing editorial discussions about your essay.

Navigation: Please be sure to describe the process by which researchers go about finding material in the database(s) through searching, browsing or other means and any special tools offered by the publisher. Please note quirks and make suggestions about how to ameliorate any difficulties you discover in the search and discovery processes.

Important: Please make screenshots as necessary and appropriate to illustrate the points that you decide to address regarding navigation.

Bringing errors to the attention of the publisher: We will need to decide how and who should make contact with database publishers detailing any errors discovered in the review process. However we decide to proceed, it is important that reviewers keep careful notations about any such errors encountered while using databases under review.

Illustrations: The essays should take advantage of the digital affordances of the web to help users visualize the attributes and limitations of the database(s). Illustrations (most likely screen shots) should be included wherever possible. We will send instructions about what image formats and dpi levels you should adhere to later on.

Citations: Unless it is completely obvious, please address any idiosyncrasies regarding the way the database should be cited. Are there archive numbers, DOIs, permanent URLs?

Hyperlinks: We expect that authors will incorporate "hot" links within the body of their writing. However, please make note of these considerations for links and stability that have been taken from the *Kairos* journal website, which can be found here: http://kairos.technorhetoric.net/styleguide.html

- All links should contribute to the possible meanings and readings of the texts. Linking for the sake of linking is
 discouraged (e.g., external links in-text to outside sources is usually discouraged in favor of links in the works
 cited; internal linking to the works cited is discouraged unless a text specifically requires it, and then backnavigation must also be provided).
- Authors should attempt to make clear where links are going so that readers may make informed navigational decisions. This can often be done by linking from descriptive phrases rather than individual words.
- Links to external nodes should point, to the best of the author's knowledge, to stable sites and resources. ...We must strive to make all links as current and accurate as possible. Authors might consider contacting the authors of pages they link to in hopes of determining such stability.
- Offsite/external links should open in a new browser window.
- Do not link terminal punctuation.

Bibliography: Every essay must include a formal, complete bibliography. The house style is Chicago 16. Additionally, because we work with a governing ethos of making scholarly work as accessible as possible, we strongly encourage, if possible, citations to open access (OA) version of works. Citations should include URLs whenever possible, with a default to OA versions. (See the section on "Hyperlinks" in this document.) The managing editor is **happy** to help you locate OA versions.

Conflicts and editorial independence: Please let the managing editor and publisher know if you have any current or prior relationships with any of the entities or institutions you are writing about. For example, if you've been on an advisory board, have received or anticipate receiving travel funding or honoraria, please let us know so we can discuss whether there are relationships that require disclosure to readers. Relatedly, while it's fine and even desirable to reach out to entities that you are covering for fact-checking purposes, please do not share the contents of your review with the entity that you are covering prior to publication without discussing it first with the publisher and editor.

Appendix B

Beyond Citation data structure (initial-December 2016)

type	field name (server side)	field name external (web side)	field name exposed externally	Entry required	field length
id	id		n	У	TBD
text	name_database		n	y	short
text (narrative)	summary		n	y	unlimited
text (narrative)	bc_overview	Date range	У	y	unlimited
text (not numeric string)	date_range		toggle content/null	n	short
text	publisher_name	Publisher	y	y	short
text (URL)	link_publisher_aboutpage	Publisher About	y	y	short
		page			
all applicable from these categories: books, journals, newspapers, images, ephemera, indexes, artifacts, media	type_object	Object type	y	у	NA
text freeform or no entry	geographic_location_original_materials	Location of original materials	toggle content/null	n	short
text freeform or no entry	geographic_location_subject	Location of subject matter	toggle content/null	n	short
Boolean y/n	image_exportable	Exportable image	toggle content/null	n	Boolean
Boolean y/n	facsimile_image	Facsimile image	toggle content/null	n	Boolean
Boolean y/n	full_text_searchable	Full text searchable	y	y	Boolean
text	link_titles_list	Titles list link	toggle content/null	n	medium
text	original_catalog	Original catalog	toggle content/null	n	short
y/n or some	original_microfilm	Original microfilm	toggle content/null	n	short

Appendix B

text (narrative)	original_sources	Original sources	toggle content/null	n	unlimited
text (narrative)	history	History/Provenance	y	y	unlimited
text (narrative)	third_party_reviews	Reviews	toggle content/null	n	unlimited
text (URL)	link_worldcat	To see the library closest to you has access:	toggle content/null	n	short
text (narrative)	access	Access	y	у	unlimited
text (narrative)	ill_conditions	InterLibrary Loan Conditions	toggle content/null	n	unlimited
text (narrative)	info_from_publisher	Info from Publisher	y	n	unlimited
text (narrative)	conversations	Conversations	y	y	unlimited
text (narrative)	citing	Citing	y	y	unlimited
date	bc_overview_date		n	y	mm-dd-yyyy
date (automated)	date_entry_last_updated	Last updated	y	y	mm-dd-yyyy
text	bc_editor_entry		n	y	short